PLOS | ONE

RESEARCH ARTICLE

The Quality of the Evidence According to GRADE Is Predominantly Low or Very Low in Oral Health Systematic Reviews

Nikolaos Pandis[1,4]*, Padhraig S. Fleming[2], Helen Worthington[3], Georgia Salanti[1]

1 Department of Hygiene and Epidemiology, Medical School, University of Ioannina, Ioannina, Greece, **2** Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Turner St., London E1 2 AD, United Kingdom, **3** Cochrane Oral Health Group, School of Dentistry, The University of Manchester, Coupland 3 Building, Oxford Road, Manchester, M13 9PL, United Kingdom, **4** Department of Orthodontics and Dentofacial Orthopedics, Dental SChool/Medical Faculty, University of Bern, Bern, Switzerland

* npandis@yahoo.com

Citation: Pandis N, Fleming PS, Worthington H, Salanti G (2015) The Quality of the Evidence According to GRADE Is Predominantly Low or Very Low in Oral Health Systematic Reviews. PLoS ONE 10(7): e0131644. doi:10.1371/journal.pone.0131644

Editor: Christian Gluud, Copenhagen University Hospital, DENMARK

Received: February 16, 2015

Accepted: June 5, 2015

Published: July 10, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by a grant of the Papavramidès Foundation, Bern, Switzerland to the first author. Additionally, Georgia Salanti received funding from the European Research Council (ERC Starting Grant IMMA 260559). These funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Objectives

The main objective was to assess the credibility of the evidence using Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) in oral health systematic reviews on the Cochrane Database of Systematic Reviews (CDSR) and elsewhere.

Study Design and Setting

Systematic Reviews or meta-analyses (January 2008-December 2013) from 14 high impact general dental and specialty dental journals and the Cochrane Database of Systematic Reviews were screened for meta-analyses. Data was collected at the systematic review, meta-analysis and trial level. Two reviewers applied and agreed on the GRADE rating for the selected meta-analyses.

Results

From the 510 systematic reviews initially identified 91 reviews (41 Cochrane and 50 non-Cochrane) were eligible for inclusion. The quality of evidence was high in 2% and moderate in 18% of the included meta-analyses with no difference between Cochrane and non-Cochrane reviews, journal impact factor or year of publication. The most common domains prompting downgrading of the evidence were study limitations (risk of bias) and imprecision (risk of play of chance).

Conclusion

The quality of the evidence in oral health assessed using GRADE is predominantly low or very low suggesting a pressing need for more randomised clinical trials and other studies of higher quality in order to inform clinical decisions thereby reducing the risk of instituting potentially ineffective and/or harmful therapies.

Introduction

Systematic reviews aim to assimilate high-quality evidence on the area of interest in a systematic, transparent, and unbiased manner leading to qualitative or quantitative synthesis. Quantitative synthesis can produce a more precise estimate on the efficacy and safety of a therapy, and can reconcile misunderstandings and controversies, and form the basis for future trials [1]. Healthcare practitioners should seek high quality evidence when searching for answers to clinical questions in relation to the effectiveness or otherwise of a proposed intervention. Several meta-epidemiological studies have been published on the quality of systematic reviews in oral health, with evidence to suggest both reporting and methodological deficiencies common to systematic reviews within a range of specialty areas [2–5]; these findings have also mirrored analogous research studies in other biomedical areas [6,7]. However, the overall quality of the existing evidence in oral health has not yet been assessed. Moreover, there is increasing concern of a gulf between research evidence and its clinical applicability. Patients are not typically conversant in the scientific publications but are exercised by the implications of research findings on their lives and wellbeing. It has become evident that a system capable of simultaneously assessing the quality of the evidence, balancing benefits and harms, while accounting for patient preferences and aiding clear treatment recommendations is imperative [8]. This approach would resonate both within medicine and in dentistry where an increasing number and quality of systematic reviews are published.

Several groups have proposed complex methods for evaluating and translating evidence into clinical practice; many of these have been somewhat confusing and impractical [9]. The GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) initiative, however, has become an accepted approach for assessing the evidence and consequently making recommendations [8]. The GRADE approach is predicated on a precise clinical question, with consideration of all important outcomes within a systematic review prioritizing them based on their relative importance [10]. Subsequently, the existing quality of the evidence for an intervention and a particular outcome from a systematic review is assessed based on a specific protocol and graded as high, moderate, low or very low (Table 1) [11].

While, precursors relied heavily on the overall study design (randomized compared to non-randomized studies) to evaluate a body of evidence, study design remains important within the GRADE assessment, although not the sole arbiter of the quality of evidence. Randomized clinical trials provide a higher quality of evidence compared to observational studies; however, good quality observational studies are more valuable than uncontrolled case series designs [12]. Randomized clinical trials (RCTs) are initially rated as high quality but may be downgraded after accounting for study limitations, indirectness, inconsistency, imprecision, and publication bias. Observational studies start at low quality but may be upgraded in the presence of a large magnitude of effect after potential confounding and dose-response effects have been considered. In the GRADE system, although expert opinion does not command a quality of

Table 1. Categories of quality of evidence according to GRADE. A single downgrade (for randomized clinical trials) or upgrade (for non-randomized studies) corresponds to one level change in the ranking.

Rank	
High ++++	Further research is very unlikely to change our confidence in the estimate of effect
Moderate +++	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low ++	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very Low +	Any estimate of effect is very uncertain

doi:10.1371/journal.pone.0131644.t001

evidence score, it is considered critical in interpreting, combining and placing the available evidence in the correct context [12].

The GRADE approach has been endorsed by a large number of societies and institutions including The Cochrane Collaboration [13] and has been used in various fields of medicine such as allergies, heart disease, oncology, endocrinology and respiratory and critical care medicine [14–17]. GRADE has also been used in two small empirical studies to appraise the quality of evidence to support medical interventions in two medical areas [18,19]. While GRADE is applicable within the oral health field [20,21], there are no previous reports of an overall assessment of the quality of evidence in this area. Therefore, the aim of this study is to assess the quality of the evidence in contemporary oral health reviews using GRADE, and to assess possible differences in the quality of the evidence between Cochrane systematic reviews and non-Cochrane reviews.

Methods

The following pre-specified inclusion criteria were applied for the systematic reviews and meta-analyses:

Eligible systematic reviews or meta-analyses for inclusion in this study included at least one meta-analysis of at least two original studies; only one meta-analysis per systematic review was considered.

The selected meta-analysis was that reporting on the primary outcome or the first or most important reported outcome if the primary outcome was not specifically outlined.

If meta-analysis of more than one primary outcome was available, the meta-analysis including the largest number of trials was selected.

The following pre-specified exclusion criteria were applied:

- Systematic reviews with forest plots without a pooled estimate.

- Duplicate publications, laboratory studies and reviews of animal studies.

-Quantitative synthesis within single arms either due to absence of a control or due to analysis of before-after measurements.

- Reviews including meta-analysis, which included the same studies multiple times without explanation on whether subgroups were mutually exclusive.

- Systematic reviews or reviews using network meta-analysis.

- Diagnostic test accuracy reviews or reviews on prognostic factors.

Systematic reviews or meta-analyses published from January 2008 (year of GRADE adoption by the Cochrane Collaboration) until the end of 2013 were retrieved from the Oral Health Group (OHG) of the Cochrane Database of Systematic Reviews (CDSR) and by hand searching of 14 general and specialty dental journals (January 2008-December 2013) with the highest impact factor (IF) in 2012 (Table 2).

The titles and abstracts were initially read by one investigator (NP) and all full-text articles were retrieved and screened for inclusion. A second screening of the full reports was undertaken by one investigator (NP) resulting in exclusion of the reviews from further analyses based on the pre-specified exclusion criteria. Information was collected from all selected review articles at the review, meta-analysis and trial level (S1 Table).

The conclusions of selected meta-analyses were assessed using the GRADE approach by the first author, who was familiar with GRADE. A second author (PSF) verified the ratings; any disagreements were reconciled after discussion. The selected meta-analyses were assessed in

Table 2. Frequencies and percentages of included systematic reviews per journal type and journal impact factors.

		Impact factor	N (%)
Journal	American Journal of Orthodontics and Dentofacial Orthopedics (AJODO)	1.46	5 (5.5%)
	Caries Research (CR)	2.51	1 (1.1%)
	COCHRANE Database of Systematic Reviews: Oral Health Group (CDSR-OHG)	5.79	41 (45.1%)
	Clinical Implant Dentistry and Related Research (CIDRR)	3.82	4 (4.4%)
	Clinical Oral Investigations (COI)	2.20	3 (3.3%)
	Clinical Oral Implant Research (COIR)	3.43	5 (5.5%)
	International Journal of Prosthetic Dentistry (IJPD)	1.63	1 (1.1%)
	Journal of Clinical periodontology (JCP)	3.69	7 (7.7%)
	Journal of Dentistry (JD)	3.20	4 (4.4%)
	Journal of Dental Research (JDR)	3.83	4 (4.4%)
	Journal of Endodontics (JOE)	2.93	3 (3.3%)
	Journal of Oral and maxillofacial Surgery (JOMS)	1.52	6 (6.6%)
	Journal of Periodontology (JP)	2.40	7 (7.7%)
Total			91 (100%)

doi:10.1371/journal.pone.0131644.t002

relation to the quality of the evidence scored in the 5 domains specified within GRADE: Limitations in study design and/or execution (risk of bias) [12], inconsistency of results [22], indirectness of evidence [23], imprecision of results [24], and publication bias [25]. The overall GRADE rating results based on 4 levels (high, moderate, low and very low) are shown in Table 1. A more detailed description of the GRADE process is shown in the appendix (S1 File and S2 Table). The GRADE judgment for Cochrane reviews was re-assessed independently of the findings of the original review authors.

Data Analysis

The objectives of the statistical analyses were to tabulate frequency distributions of specific characteristics in the dental systematic review sample at the review, and meta-analysis levels together with the frequency distributions of the quality of the evidence in relation to GRADE for Cochrane systematic reviews and non-Cochrane reviews (or meta-analyses). Associations between systematic review/meta-analysis characteristics and GRADE assessment were also considered. The four-level GRADE rating (high, moderate, low, very low) was converted into a binary variable (high/moderate and low/very low) in order to fit a logistic regression model to assess potential associations between GRADE rating, impact factor and publication year. GRADE was the dependent variable and impact factor and publication year were the examined predictors. All the analyses were performed with Stata statistical software version 13.1 (Stata Corporation, College Station, Texas, USA).

Results

From the 510 reviews initially considered for inclusion in the study, 91 were included in the final assessment (Fig 1). The frequencies of reviews per journal are outlined in Table 2. Fifty reviews were included from the selected dental journals and 41 systematic reviews from The Cochrane Database for Systematic Reviews; a variety of conditions, interventions and outcomes were considered. The number of published systematic reviews with meta-analyses increased over time (S1 Fig). Significant differences were found between Cochrane and non-Cochrane reviews in respect of the region of authorship, involvement of a methodologist, number of collaborating centers, inclusion of GRADE assessment and assessment of harms in the

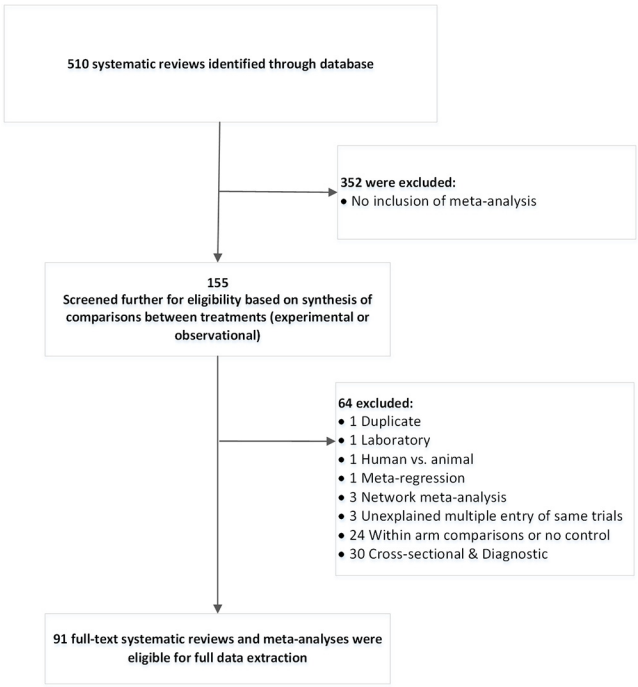

Fig 1. Systematic review selection flow diagram.

doi:10.1371/journal.pone.0131644.g001

outcome category. The Cochrane compared to non-Cochrane reviews are more likely to originate in Europe (OR: 14.62, 95% CI: 1.76, 121.19, p = 0.01) or Asia (OR: 1.43, 95% CI: 0.11, 18.00, p = 0.78) compared to Americas, to involve authors across multiple centers (OR: 5.26, 95% CI: 1.97, 14.09, p = 0.001), to include a methodologist (OR = 73.76, 95% CI:9.79, 555.93, p<0.001) and a GRADE assessment (OR = 12.71, 95% CI:4.17, 38.69, p<0.001), and to consider at least one harm in the outcomes (OR = 2.66, 95% CI: 1.13, 6.27, p = 0.03) (S3 Table).

The Cochrane systematic reviews included only randomized clinical trials but the non-Cochrane reviews included both randomized and non-randomized studies and occasionally a combination of both designs. Authors of Cochrane reviews were more likely to account for clustering effects (OR: 15.62, 95% CI: 2.81, 86.76, p = 0.002) often encountered in dentistry either in the analysis or discussion and to correctly analyze paired data (OR: 11.25, 95% CI: 11.05, 541.20, p = 0.02) compared to non-Cochrane reviews (S4 Table).

A variety of approaches were used for the assessment of the methodological quality with the Cochrane risk of bias tool used in 62 out of the 91 reviews. In some reviews, a combination of approaches was used and in some instances incorrectly reporting guidelines were utilized for this purpose.

The GRADE assessment indicated that only 2% of the evidence belongs to the high and 18% to the moderate category with the remaining reviews at the low/very low level (Table 3). The distribution of the overall GRADE ratings and ratings for each GRADE domain were similar for Cochrane and non-Cochrane reviews (Table 3).

The most commonly downgraded domains were those for study limitations (risk of bias) and imprecision. The frequency of downgrading by 2 levels (very serious) was higher in the non-Cochrane reviews (46%) compared to the Cochrane reviews (22%); the former include also non-randomized studies. The inconsistency domain received similar ratings for non-

Table 3. Frequencies and percentages of overall and per domain GRADE rating for Cochrane and non-Cochrane meta-analyses.

	Non- Cochrane Review	Cochrane Systematic Review	Total	p-value #, ##
	N (%)	N (%)	N (%)	
GRADE rating				
High	2 (4%)	0 (0%)	2 (2%)	ns
Moderate	9 (18%)	11 (27%)	20 (22%)	
Low	24(48%)	19 (47%)	43 (47%)	
Very Low	15(30%)	11 (27%)	26 (29%)	
Domains				
Study limitations/Risk of Bias				
No	2 (4%)	2 (5%)	4 (4%)	ns
Serious	25(50%)	30 (73%)	55 (60%)	
Very Serious	23(46%)	9 (22%)	32 (35%)	
Inconsistency				
No	34(72%)	31(76%)	65 (71%)	ns
Serious	15(26%)	10(24%)	25 (27%)	
Very Serious	1(2%)	0(0%)	1 (1%)	
Indirectness				
No	50(100%)	38 (93%)	88 (97%)	ns
Serious	0 (0%)	3 (7%)	3 (3%)	
Very serious	0 (0%)	0 (0%)	0 (0%)	
Imprecision				
No	19(38%)	19 (46%)	38 (42%)	ns
Serious	23(46%)	18 (44%)	41 (45%)	
Very Serious	8 (16%)	4 (10%)	12 (13%)	
Publication bias				
Suspected	8 (16%)	0 (0%)	8 (9%)	**
Unsuspected	42(84%)	41 (100%)	83 (91%)	
GRADE rating as reported in the review				
High	2 (4%)	0 (0%)	2 (2%)	
Moderate	1 (2%)	9 (22%)	10 (11%)	
Low	0 (0%)	10 (24%)	10 (11%)	
Very Low	2 (4%)	3 (8%)	5 (6%)	
Not reported	45(10%)	19 (46%)	64 (70%)	
Total	50 (100%)	41 (100%)	91(100%)	

ns = non-significant

*** Significant at 0.01*

#Pearson X^2 test or Fisher's exact test

P-value based on High/Moderate vs Low/Very Low for GRADE and on No vs Serious/Very serious for the GRADE domain comparisons

doi:10.1371/journal.pone.0131644.t003

Cochrane and Cochrane reviews in agreement with the reported statistical heterogeneity represented by I^2 (S2 Fig).

A GRADE assessment was included in 5/50 (10%) of the non-Cochrane and in 22/41 (54%) of the Cochrane reviews, respectively. Only three discrepancies were observed among the 27 included in the systematic reviews between the evidence rating found by the initial review authors and our re-analysis. One review published [26] in the Journal of Clinical Periodontology was based on a previous Cochrane systematic review which did not include a GRADE

assessment [27]. Logistic regression analysis indicated that neither the journal impact factor (OR = 0.92, 95% CI: 0.68, 1.25, p = 0.61) nor the year of publication (OR: 1.17, 95% CI: 0.82, 1.67, p = 0.40) are important predictors for the quality of the evidence (Reference group: low/very low evidence).

Discussion

This cross-sectional meta-epidemiologic study is the first attempt to evaluate the evidence across the oral health field and has exposed that only 2% of the existing evidence derived from reviews has a high level of evidence and only 18% was of a moderate level of evidence in relation to primary outcomes with no difference between Cochrane and non-Cochrane reviews. This finding indicates that 80% of the evidence considered, much of which will form the basis of clinical recommendations, is at the low or very low level. This finding is alarming indicating the need to undertake high quality randomized clinical trials in order to reduce the uncertainty around therapies in the oral health field. Empirical studies in oral health and medicine in general using GRADE to assess the quality of evidence are sparse [18,19]. Our study included 91 reviews and covered the entire field of oral health systematic reviews and meta-analyses; it is likely that the quality of the evidence may vary across, between and within healthcare specialties. Although, GRADE has been used to assess the quality of evidence within specific systematic reviews in several fields, we were able to identify only two small empirical studies across the range of biomedical fields where GRADE was applied in meta-epidemiological studies on a body of systematic reviews to assess the quality of the evidence within a particular specialty. The scope of those studies was limited to hyperbaric oxygen therapy indications [19] and nonsurgical treatment of stress urinary incontinence [18]. In the analysis of hyperbaric oxygen treatment[19], the quality of the evidence ranged from very low to high depending on the indication with key modifiers identified as risk of bias, imprecision and large effect for observational studies. Similarities with our study include reasons for downgrading the quality of the evidence based on risk of bias and imprecision. The review on nonsurgical treatment of stress urinary incontinence involved assessment of 13 reviews finding that the quality of the evidence ranged from low to high depending on the intervention [18].

Cochrane reviews are considered to be particularly rigorous and are conducted according to strict criteria and usually include only randomized clinical trials for assessment of benefits and harms and only use quasi-randomized studies and observational studies for the assessment of harms [2]. In the present study, the non-Cochrane reviews had a relatively small proportion of non-randomized studies; however, the quality of evidence did not differ among the groups suggesting the quality of studies populating both Cochrane and non-Cochrane reviews to be lacking. The two most frequently downgraded domains were risk of bias (study limitations) and imprecision. For the risk of bias (study limitations) domain lack of allocation concealment, lack of blinding, losses to follow-up and selective reporting commonly led to a downgrade [28]. In relation to the imprecision domain, small sample size and wide confidence intervals which include both benefit and harm are likely to prompt a downgrade [24]. Specifically, for binary outcomes, if the 95% confidence intervals ranged from unimportant to important benefits and from unimportant to important harms, the quality of evidence was downgraded when appreciable benefit or harm are of the order of 25% relative risk reduction or increase. For continuous outcomes, if the 95% confidence intervals included no effect and upper and lower confidence interval bounds crossed the minimal important clinical difference (MID) for harm or benefit downgrade was considered appropriate. If the MID was not known scores were downgraded if the upper or lower confidence limit crossed an effect size of 0.5 in either direction [24]. Downgrade for these reasons, in particular, were common as the number and size of

the included studies was usually small. This finding was consistent with a recent study evaluating the quality of evidence in hyperbaric oxygen therapy [19]. The latter review, however, was also based on a relatively limited subset of 17 reviews, with the majority of primary studies (75%) being non-randomized studies.

In relation to the characteristics of the reviews, Cochrane reviews were more likely to be published by European authors, to involve authors across multiple centers, to involve a methodologist, and to consider at least a single harm in the outcomes than was the case for non-Cochrane reviews. However, the strict methodological and reporting criteria and the involvement of a methodologist did not result in higher quality of evidence. This is logical as, while Cochrane reviews are likely to be conducted and reported to a higher standard, published systematic reviews are based on studies from the same pool, regardless of the effect of methodological quality of the review itself, review authorship and associated expertise. This finding does not suggest that Cochrane systematic reviews are redundant with a level of commensurate with non-Cochrane reviews; Cochrane reviews are rigorous and replicable and this should be the standard moving forward.

The systematic reviews included a median of 5 meta-analyses, each comprising of a median of 5 studies. These figures are in keeping with a large survey of the Cochrane Database of Systematic Reviews involving analysis of 2,321 systematic reviews which highlighted a median of 6 meta-analyses per review and a median of 3 studies per meta-analysis [29]. A similar previous review [30] of dental systematic reviews reported a median of 9 studies were included in the largest meta-analysis within dental systematic reviews involving 9 dental specialties, with the largest meta-analysis having no more than 4 studies in 19% of reviews. Similarly, the largest meta-analysis involved a median number of just two randomized clinical trials, although that review referred back to systematic reviews from as long ago as 1991 when randomized clinical trials were considerably less prevalent in oral health than is now the case.

In oral health, due to the fact that multiple matched or unmatched sites can receive the intervention of interest, clustering effects are common. If this is handled improperly during the analysis, significant results may arise which are not genuine [31–33]. Additionally, a common design which uses matching called split-mouth studies requires consideration of the within patient correlations during the synthesis. In split-mouth designs, clustering effects can also exist when both interventions are applied within the same patient and in the presence of multiple sites per treatment arm as is the case with teeth [33, 34]. Studies with clustering effects, paired data or a mixture of these were handled more appropriately in the Cochrane systematic reviews than in the non-Cochrane meta-analyses.

The Cochrane Collaboration adopted GRADE in 2008; this may explain the inclusion of a GRADE assessment more frequently in the Cochrane Database of Systematic Reviews with 22/41 (82%) Cochrane systematic reviews including a GRADE assessment versus 5/50 (10%) among non-Cochrane reviews. Therefore, while Cochrane systematic reviews follow strict methodology, it is evident that deviations from the pre-specified ideal are possible. We conducted our own assessment of GRADE highlighting a discrepancy in only 11% of occasions with the rating of the systematic review authors confirming the reliability of the GRADE approach when assessing the evidence [35]. Other investigators found larger inconsistencies during the application of GRADE [36], although the level of familiarity of these investigators with GRADE is unclear. In the present study one of the investigators who assessed the quality of the evidence had attended a GRADE course and both GRADE assessors have implemented GRADE previously in systematic reviews.

Older systematic reviews, time from search to publication and delays in systematic review updates may lead to important differences in included studies compared to more recent reviews [37]. Publication year, however, was not found to have a bearing on the quality of the

evidence in the present study. Similarly, journal impact factor showed no association with the quality of the evidence. There is some evidence that systematic reviews published in higher impact medical journals are of higher methodological quality [38]; however, this does not necessarily translate into higher quality of evidence. Systematic review methodological quality should not be confused with the quality of the evidence, although more detailed and broader searches in particular may lead to the identification of more eligible studies and potentially lead to higher quality of evidence.

The present study did not include a cross-section of all oral health reviews but focused on the Cochrane Database of Systematic Reviews and dental journals with the highest impact factor, which may have potentially influenced the results, although the pool of potential candidate studies for inclusion is similar. Nevertheless, the observed findings are likely to represent a best-case scenario. The inclusion of the highest impact factor dental journals possibly explains the lack of association between impact factor and evidence quality. The several steps of the GRADE assessment may introduce an element of inconsistency in the ratings; it is, therefore, important that the complete picture is considered along with the judgment made. Rating of study limitations for non-randomized studies can be problematic especially when the information provided in the review is limited. A particularly challenging area to rate is publication bias as it is difficult to detect exclusion of eligible studies. Both the description of the literature search, reference to grey literature, trial registries and reference lists of included studies during the assessment of publication bias were considered to reach this decision. Statistical assessment of publication bias was typically impossible due to the small number of included studies in the meta-analyses [39]. It is of interest to note that the Agency for Healthcare Research and Quality Evidence based Practice Center Program considers publication bias to be an optional domain [40]. Finally, evaluation of imprecision can be often challenging. The GRADE approach suggests the use of confidence intervals since dichotomous decisions based on statistical significance are problematic [24]. Alternative approaches that can be used to decide upon the conclusiveness of the evidence have been proposed more recently [41].

In the present study, one author performed the initial screening of the eligible systematic reviews and one author made the GRADE rating with scores verified by a second author. In a previous study [36] implementation agreement ranged from low to high depending on the domain and concluded that both training in GRADE and clinical expertise are instrumental in improving consistency in its use. GRADE permits judgment in a methodical and transparent way [42] with inconsistencies relating to the type and number of outcomes considered during the assessment [36]. The approach favors an overall rating which may be considered as a continuum hinging on expertise and judgment resulting in a judgment where the overall may not be the sum of all parts [42]. Nevertheless, as an incidental finding, significant agreement was observed between the ratings made in the constituent reviews and in the present cross-sectional study. For the purposes of assessing the association between GRADE rating and publication year and impact factor, and because data was thin the 4-level scale was converted to two levels. We understand that GRADE uses 4 levels for a purpose and distilling the conclusions into a 2-level scale has limitations. In this instance, however, the 2-level presentation helped in communicating the results especially when applying GRADE for recommendations.

This study dealt with only one outcome for which meta-analyses were available. A more comprehensive approach could have involved consideration of several or all outcomes and could have included evidence from qualitative synthesis. However, this would have been very difficult to implement given the large number of systematic reviews. Furthermore, we feel that inclusion of only the primary or first outcome over a wide range of oral health systematic reviews is likely to be a good proxy of the quality of the evidence in the field of oral health research.

In conclusion, only a small proportion (20%) of the studies assessing interventions in oral health was of moderate (18%) or high (2%) quality according to GRADE. The most common domains provoking downgrading of the evidence were risk of bias (study limitations) and imprecision indicating the need for larger and higher quality randomized clinical trials to inform clinical decisions. The lack of robust evidence underpinning dental procedures should prompt a concerted drive to conduct funded, high quality clinical trials in order to improve the quality of the evidence for accepted but often unproven procedures.

Supporting Information

S1 Fig. Number of included non-Cochrane and Cochrane reviews by year of publications.
(TIF)

S2 Fig. Heterogeneity (I-square %) for non-Cochrane and Cochrane included meta-analyses.
(TIF)

S1 File. Detailed description of assessing the quality of the evidence per GRADE domain.
(DOCX)

S1 Table. Collected information per systematic review at review, meta-analysis, trial level and GRADE assessment.
(DOCX)

S2 Table. GRADE assessment at the selected meta-analysis level. Mixed indicates inclusion of both randomized and non-randomised studies in the meta-analysis.
(DOCX)

S3 Table. Frequencies and percentages of included review characteristics for Cochrane and non-Cochrane reviews at review level.
(DOCX)

S4 Table. Frequencies and percentages of characteristics of selected meta-analyses for detailed assessment for 41 Cochrane and 50 not Cochrane reviews at meta-analysis level.
(DOCX)

Acknowledgments

We would like to thank Ms Nancy Santesso from the GRADE group for helpful comments during the preparation of this manuscript.

Author Contributions

Conceived and designed the experiments: NP HW GS. Performed the experiments: NP PSF. Analyzed the data: NP GS. Contributed reagents/materials/analysis tools: NP PSF. Wrote the paper: NP PSF HW GS.

References

1. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med. 2009; 151: W65–94. PMID: 19622512
2. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: a quality paradigm? Eur J Orthod. 2013; 35: 244–248. doi: 10.1093/ejo/cjs016 PMID: 22510325

3. Seehra J, Fleming PS, Polychronopoulou A, Pandis N. Reporting completeness of abstracts of systematic reviews published in leading dental specialty journals. Eur J Oral Sci. 2013; 121: 57–62. doi: 10.1111/eos.12027 PMID: 23489893
4. Kiriakou J, Pandis N, Fleming PS, Madianos P, Polychronopoulou A. Reporting quality of systematic review abstracts in leading oral implantology journals. J Dent. 2013; 41: 1181–1187. doi: 10.1016/j.jdent.2013.09.006 PMID: 24075952
5. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. Angle Orthod. 2013; 83: 158–163. doi: 10.2319/032612-251.1 PMID: 22720835
6. Martel G, Duhaime S, Barkun JS, Boushey RP, Ramsay CR, Fergusson DA. The quality of research synthesis in surgery: the case of laparoscopic surgery for colorectal cancer. Syst Rev. 2012; 1: 14. doi: 10.1186/2046-4053-1-14 PMID: 22588035
7. Brito JP, Tsapas A, Griebeler ML, Wang Z, Prutsky GJ, Domecq JP, et al. Systematic reviews supporting practice guideline recommendations lack protection against bias. J Clin Epidemiol. 2013; 66: 633–638. doi: 10.1016/j.jclinepi.2013.01.008 PMID: 23510557
8. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336: 924–926. doi: 10.1136/bmj.39489.470347.AD PMID: 18436948
9. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res. 2004; 4: 38. doi: 10.1186/1472-6963-4-38 PMID: 15615589
10. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011; 64: 395–400. doi: 10.1016/j.jclinepi.2010.09.012 PMID: 21194891
11. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64: 383–394. doi: 10.1016/j.jclinepi.2010.04.026 PMID: 21195583
12. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011; 64: 401–406. doi: 10.1016/j.jclinepi.2010.07.015 PMID: 21208779
13. http://www.gradeworkinggroup.org/society/index.htm [Internet].
14. De Palma R, Liberati A, Ciccone G, Bandieri E, Belfiglio M, Ceccarelli M, et al. Developing clinical recommendations for breast, colorectal, and lung cancer adjuvant treatments using the GRADE system: a study from the Programma Ricerca e Innovazione Emilia Romagna Oncology Research Group. J Clin Oncol Off J Am Soc Clin Oncol. 2008; 26: 1033–1039. doi: 10.1200/JCO.2007.12.1608
15. Swiglo BA, Murad MH, Schünemann HJ, Kunz R, Vigersky RA, Guyatt GH, et al. A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the grading of recommendations, assessment, development, and evaluation system. J Clin Endocrinol Metab. 2008; 93: 666–673. doi: 10.1210/jc.2007-1907 PMID: 18171699
16. Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force. Chest. 2006; 129: 174–181. doi: 10.1378/chest.129.1.174 PMID: 16424429
17. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy. 2009; 64: 669–677. doi: 10.1111/j.1398-9995.2009.01973.x PMID: 19210357
18. Latthe P, Foon R, Khan K. Nonsurgical treatment of stress urinary incontinence (SUI): grading of evidence in systematic reviews. BJOG Int J Obstet Gynaecol. 2008; 115: 435–444. doi: 10.1111/j.1471-0528.2007.01629.x
19. Murad MH, Altayar O, Bennett M, Wei JC, Claus PL, Asi N, et al. Using GRADE for evaluating the quality of evidence in hyperbaric oxygen therapy clarifies evidence limitations. J Clin Epidemiol. 2014; 67: 65–72. doi: 10.1016/j.jclinepi.2013.08.004 PMID: 24189086
20. Faggion CM Jr. The shortened dental arch revisited: from evidence to recommendations by the use of the GRADE approach. J Oral Rehabil. 2011; 38: 940–949. doi: 10.1111/j.1365-2842.2011.02230.x PMID: 21707696
21. Faggion CM Jr. Grading the quality of evidence and the strength of recommendations in clinical dentistry: a critical review of 2 prominent approaches. J Evid-Based Dent Pract. 2010; 10: 78–85. doi: 10.1016/j.jebdp.2010.01.001 PMID: 20466314

22. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. 2011; 64: 1294–1302. doi: 10.1016/j.jclinepi.2011.03.017 PMID: 21803546

23. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol. 2011; 64: 1303–1310. doi: 10.1016/j.jclinepi.2011.04.014 PMID: 21802903

24. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. J Clin Epidemiol. 2011; 64: 1283–1293. doi: 10.1016/j.jclinepi.2011.01.012 PMID: 21839614

25. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. J Clin Epidemiol. 2011; 64: 1277–1282. doi: 10.1016/j.jclinepi.2011.01.011 PMID: 21802904

26. Eberhard J, Jervøe-Storm P-M, Needleman I, Worthington H, Jepsen S. Full-mouth treatment concepts for chronic periodontitis: a systematic review. J Clin Periodontol. 2008; 35: 591–604. doi: 10.1111/j.1600-051X.2008.01239.x PMID: 18498383

27. Eberhard J, Jepsen S, Jervøe-Storm P-M, Needleman I, Worthington HV. Full-mouth disinfection for the treatment of adult chronic periodontitis. Cochrane Database Syst Rev. 2008; CD004622. doi: 10.1002/14651858.CD004622.pub2

28. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011; 64: 407–415. doi: 10.1016/j.jclinepi.2010.07.017 PMID: 21247734

29. Davey J, Turner RM, Clarke MJ, Higgins JPT. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol. 2011; 11: 160. doi: 10.1186/1471-2288-11-160 PMID: 22114982

30. Saltaji H, Cummings GG, Armijo-Olivo S, Major MP, Amin M, Major PW, et al. A descriptive analysis of oral health systematic reviews published 1991–2012: cross sectional study. PloS One. 2013; 8: e74545. doi: 10.1371/journal.pone.0074545 PMID: 24098657

31. Fleming PS, Koletsi D, Polychronopoulou A, Eliades T, Pandis N. Are clustering effects accounted for in statistical analysis in leading dental specialty journals? J Dent. 2013; 41: 265–270. doi: 10.1016/j.jdent.2012.11.012 PMID: 23201411

32. Koletsi D, Pandis N, Polychronopoulou A, Eliades T. Does published orthodontic research account for clustering effects during statistical data analysis? Eur J Orthod. 2012; 34: 287–292. doi: 10.1093/ejo/cjr122 PMID: 22015822

33. Pandis N, Walsh T, Polychronopoulou A, Eliades T. Cluster randomized clinical trials in orthodontics: design, analysis and reporting issues. Eur J Orthod. 2013; 35: 669–675. doi: 10.1093/ejo/cjs072 PMID: 23041934

34. Pandis N, Walsh T, Polychronopoulou A, Katsaros C, Eliades T. Split-mouth designs in orthodontics: an overview with applications to orthodontic clinical trials. Eur J Orthod. 2013; 35: 783–789. doi: 10.1093/ejo/cjs108 PMID: 23376899

35. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. J Clin Epidemiol. 2013; 66: 736–742.e5. doi: 10.1016/j.jclinepi.2013.02.004 PMID: 23623694

36. Hartling L, Fernandes RM, Seida J, Vandermeer B, Dryden DM. From the Trenches: A Cross-Sectional Study Applying the GRADE Tool in Systematic Reviews of Healthcare Interventions. Malaga G, editor. PLoS ONE. 2012; 7: e34697. doi: 10.1371/journal.pone.0034697 PMID: 22496843

37. Beller EM, Chen JK-H, Wang UL-H, Glasziou PP. Are systematic reviews up-to-date at the time of publication? Syst Rev. 2013; 2: 36. doi: 10.1186/2046-4053-2-36 PMID: 23714302

38. Fleming PS, Koletsi D, Seehra J, Pandis N. Systematic reviews published in higher impact clinical journals were of higher quality. J Clin Epidemiol. 2014; doi: 10.1016/j.jclinepi.2014.01.002

39. Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

40. Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions—agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol. 2010; 63: 513–523. doi: 10.1016/j.jclinepi.2009.03.009 PMID: 19595577

41. Ferreira ML, Herbert RD, Crowther MJ, Verhagen A, Sutton AJ. When is a further clinical trial justified? BMJ. 2012; 345: e5913. doi: 10.1136/bmj.e5913 PMID: 22977141

42. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2013; 66: 151–157. doi: 10.1016/j.jclinepi.2012.01.006 PMID: 22542023